STORAGE SUBSYSTEM AND STORAGE CONTROLLER

The present application is a continuation of application Serial No. 09/608,151, filed June 30, 2000, the contents of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

The present invention relates to a storage subsystem and a storage controller, both connected to host computers. More particularly, the invention relates to a storage subsystem and a storage controller adapted to provide enhanced performance and reliability.

Description of the Related Art

In recent years, storage controllers have been required to provide better performance, higher reliability and greater availability than ever before as computer systems are getting larger in scale to process data at higher speeds than ever before, 24 hours a day and 365 days a year, with data transfer interfaces also enhanced in speed. Illustratively, Japanese Patent Laid-open No. Hei 11-7359 discloses a storage controller incorporating an internal network to improve its performance.

There has been a growing need for connecting a storage controller to a plurality of host computers having multiple interfaces, as shown in Fig. 8. In such a storage controller, a host interface section comprises a host interface for addressing each different host computer. A control processor in each host interface analyzes I/O commands received from the corresponding host computer and exchanges data accordingly with a cache memory 215 over a signal line. Japanese Patent Laid-open

No. Hei 9-325905 illustratively discloses one such storage controller.

Techniques have been known recently which substitute a fibre channel interface for the SCSI (Small Computer System Interface) between a host computer and a storage controller. Illustratively, Japanese Patent Laid-open No. Hei 10-333839 discloses techniques for connecting a storage controller with a host computer using a fibre channel interface. The disclosed storage controller is designed for dedicated use with a host computer having a fibre channel interface.

SUMMARY OF THE INVENTION

The techniques disclosed in the above-cited Japanese Patent Laid-open Nos. Hei 11-7359 and Hei 9-325905 have one disadvantage: the storage controller as a whole has its performance constrained by the performance of a single control processor that handles I/O requests from host computers. Another disadvantage is that a disabled control processor will prevent host computers from using the storage controller. In particular, since today's fibre channels are capable of transferring data at speeds as high as 100 MB/s, the performance of the control processor can be an impediment to taking advantage of the high data transfer rates offered by fibre channels.

The techniques disclosed in the above-cited Japanese Patent Laid-open No. Hei 10-333839 relate to a storage controller for exclusive use with fibre channel interfaces. That is, the proposed storage controller is incapable of connecting with a host computer having a SCSI interface.

It is therefore an object of the present invention to provide a storage subsystem and a storage controller adapted to take advantage of high data transfer rates of fibre channels while offering enhanced reliability and availability.

It is another object of the present invention to provide a storage subsystem and a storage controller capable of connecting with a plurality of host computers having multiple different interfaces.

In carrying out the invention and according to one aspect thereof, there is provided a storage subsystem or a storage controller for controlling transfer of input/output data to and from a lower level storage medium drive unit in response to input/output requests received from a higher level external entity. The storage subsystem or storage controller comprises: at least one external interface controller for receiving the input/output requests from the higher level external entity in accordance with a type of interface with the higher level external entity; at least one control processor which processes the input/output requests; and a loop of fibre channel interfaces interposed between the external interface controller and the control processor so as to serve as a channel through which information is transferred therebetween.

In a preferred structure according to the invention, the interface of the external interface controller interfacing to the higher level external entity may be a fibre channel interface. In another preferred structure according to the invention, the external interface controller may be capable of interface conversion between an interface which interfaces to the higher order external entity and which is different from a fibre channel interface on the one hand, and a fibre channel interface on the other hand.

Other objects, features and advantages of the invention will become more apparent upon a reading of the following description and appended drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram of a storage subsystem practiced as an embodiment of the invention;

Fig. 2 is a block diagram of a loop 133 in the embodiment and related facilities;

Fig. 3 is a table showing a data structure of FCAL management information 113 for use with the embodiment;

Fig. 4 is a flowchart of steps performed by control processors 114 through 117 of the embodiment;

Fig. 5 is a table depicting an example of FCAL management information 113 updated when control processors stopped;

Fig. 6 is a table indicating an example of FCAL management information 113 updated when an imbalance of control processor loads was detected;

Fig. 7 is a table showing another example of FCAL management information 113 updated when an imbalance of control processor loads was detected; and

Fig. 8 is a block diagram of a conventional storage controller.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Preferred embodiments of this invention will now be described with reference to the accompanying drawings.

Fig. 1 is a block diagram of a system comprising a disk subsystem typically embodying the invention. A disk controller 107 is connected to host computers 100, 101 and 102 on the higher level side. The host computer 101 is a mainframe computer connected to the disk controller 107 through a mainframe channel. The host computer 100 is an open system computer connected to the disk controller 107

through a fibre channel interface. The host computer 102 is another open system computer connected to the disk controller 107 via a SCSI (Small Computer System Interface). The disk controller 107 is connected via loops 125 and 126 of fibre channel interfaces to drives 127, 128, 129 and 130 on the lower level side.

Host interface controllers (HIFC) 103, 104 and 105 are connected to the host computers 100, 101 and 102 respectively, as well as to a loop 133 of fibre channel interfaces. Control processors 114, 115, 116 and 117 are connected to the loop 133 on the one hand and to a common bus 118 on the other hand. The common bus 118 is connected not only to the controller processors 114 through 117 but also to a shared control memory 112, a cache memory 122, and control processors 119 and 120. The control processors 119 and 120 are connected via fibre channels 141 to drive interface controllers (DIFC) 123 and 124. The DIFCs 123 and 124 are connected to the drives 127, 128, 129 and 130 through the loops 125 and 126. The control processors 114, 115, 116 and 117 are connected to a service processor 131 by way of a signal line 132.

The HIFC 103 is an interface controller interfacing to a higher level external entity. Upon receipt of I/O commands, data and control information in the form of frames from the host computer 100, the HIFC 103 forwards what is received unmodified to one of the control processors 114 through 117 through the loop 133. On receiving data and control information in frames from any of the control processors 114 through 117 via the loop 133, the HIFC 103 transfers the data and information unmodified to the host computer 100. The HIFC 104 converts channel commands, data and control information received from the host computer 101 into fibre channel frame format for transfer to one of the control processors 114 through 117 via the loop 133. Upon receipt of data and control information in frames from any

of the control processors 114 through 117, the HIFC 104 converts the received data and information into a data format compatible with a mainframe channel interface before transferring what is converted to the host computer 101. The HIFC 105 converts I/O commands, data and control information received from the host computer 102 into fibre channel frame format for transfer to one of the control processors 114 through 117. The HIFC 105 receives data and control information in frames from any of the control processors 114 through 117, and converts the received data and information into SCSI compatible data format for transfer to the host computer 102. It is possible to connect a plurality of host computers 100, 101, 102, etc., to each of the HIFCs 103, 104 and 105.

The cache memory 122 may be accessed by all control processors 114 through 117, 119 and 120 via a bus interface of the common bus 118. When in use, the cache memory 122 temporarily accommodates data sent from the host computers 100 through 102 as well as data retrieved from the drives 127 through 130. The data in the cache memory 122 are divided into data management units called cache slots.

The shared control memory 112 may be accessed by all control processors 114 through 117, 119 and 120 via the common bus 118. This memory has regions permitting communication between the control processors, and a cache slot management table, and stores FCAL (fibre channel arbitrated loop) management information 113 for establishing frames to be received through the loop 133 by each of the control processors 114 through 117. Each of the control processors 114 through 117 references the FCAL management information 113 in the shared control memory 112 to capture a frame having a relevant address from among the frames flowing through the loop 133, and executes an I/O request designated by a received

I/O command. Upon receipt of a read command, the control processor reads the requested data if any from the cache memory 122, and sends the retrieved data to the requesting host computer through the loop 133 and via one of the HIFCs 103 through 105. If the requested data are not found in the cache memory 122, the control processor in question sends an I/O request to the control processors 119 and 120. Upon receipt of a write command, one of the control processors 114 through 117 writes target write data to a cache slot in the cache memory 122 and sends an I/O request to the control processors 119 and 120.

The control processors 119 and 120 receive an I/O request from one of the control processors 114 through 117. If a read command is received, the control processors 119 and 120 read the requested data from the drives 127 through 130 and write the retrieved data to a cache slot in the cache memory 122. In the case of a write command, the control processors 119 and 120 write the relevant data from the cache memory 122 to the drives 127 through 130. Fig. 2 is a block diagram of the loop 133 interposed between the HIFCs 103 through 106 on the one hand and the control processors 114 through 117 on the other hand, along with facilities associated with the loop 133.

The loop 133 has port bypass circuits (PBC) 108, 109, 110 and 111 constituting what is known as a hub structure. The PBCs 108 through 111 are a one-input n-output electronic switch each. As illustrated, the PBCs 108 through 111 are connected to the HIFCs 103 through 106 and to the control processors 114 through 117. Interconnections are provided between the PBCs 108 and 111 as well as between the PBCs 109 and 110. In this embodiment, the PBCs 108 through 111 serve as a one-input two-output switch each. Feeding a suitable input signal to the PBC arrangement makes it possible to limit the number of output paths. Fiber

controllers (FC) 151 disposed upstream of the control processors 114 through 117 recognize destination addresses of frames sent through the loop 133, capture a frame having a predetermined destination address, and transfer the captured frame to the relevant control processor connected. The fibre controllers 151 receive data and control information from the control processors 114 through 117, convert the received data and information into frame format data, and forward what is converted to the loop 133. With the HIFCs 103 through 106, FCs 151, and control processors 114 through 117 as its terminals, the loop 133 constitutes a topological loop transmission channel called a fibre channel arbitrated loop (FCAL). A fibre channel communication protocol is discussed illustratively in the published ANSI manual "FIBRE CHANNEL PHYSICAL AND SIGNALLING (FC-PH) REV. 4.3."

The PBC 108 is connected illustratively to the host computer 100 via the HIFC1O3. In this setup, the PBC 108 is connectable to the control processors 114 and 115 as well as to the PBC 111. This means that an I/O request command from the host computer 100 may be processed by the control processor 114 or 115 via the PBC 108 or by the control processor 116 or 117 via the PBC 111. Likewise, an I/O request command from the host computer 101 may be processed by the control processor 114 or 115 via the PBC 109 or by the control processor 116 or 117 via the PBC 110.

This embodiment adopts a fibre channel interface for the fibre channels 141 as well as for the loops 125 and 126. Thus the FCs 151, not shown, are in fact interposed between the control processors 119 and 120 on the one hand and the fibre channels 141 on the other hand.

Fig. 3 is a table showing a data structure of the FCAL management information 113. The FCAL management information 113 constitutes a table in which

frames to be captured by the control processors 114 through 117 via the loop 133 are set along with the range of device numbers subject to I/O processing. Entries making up the FCAL management information 113 include control processor numbers 201, AL-PAs (arbitrated loop physical addresses) 202, and LUNs (logical unit numbers) 203. A control processor number 201 is an identifier of any one of the control processors 114 through 117. An AL-PA 202 is an address assigned in the loop 133 to one of the control processors 114 through 117. A LUN 203 denotes a logical device number or a range of logical device numbers of devices whose I/O processing is carried out by a given control processor. The FCAL management information 113 may be set or canceled as instructed by the service processor 131.

Fig. 4 is a flowchart of steps performed by the control processors 114 through 117. Each of the control processors 114 through 117 periodically reads entries for the processor in question from the FCAL management information 113, and sets an AL-PA of the applicable processor to the connected FC 151. In case of a change, 13 the AL-PA is set again. The FC 151 reads AL-PAs in frames sent from the host computer 100 through the HIFC 103 and via the loop 133 (in step 301). If a given AL-PA is not found to be that of the connected control processor ("NO" in step 302), the processing is brought to an end. If an AL-PA is judged to be that of the connected control processor ("YES" in step 302), then the control processor in question is notified thereof. Given the notice, the applicable control processor (one of the processors 114 through 117) reads the frame via the FC 151 (in step S303). A check is made to see if the LUN of the I/O command in the frame falls within the range of the LUN 203 (in step 304). If the designated LUN does not fall within the range of the LUN 203, an error response is returned to the host computer 100. The control processor then effects an I/O request in accordance with the received I/O

command (in step 305).

If the I/O request is a write request, the control processors 114 through 117 receive data from the host computer 100, write the received data to a suitable cache slot in the cache memory 122, and terminate the write request processing. The slot number of the cache slot to which to write the data is computed from an LBA (logical block address) attached to the data. That memory address in the cache memory 122 which corresponds to the slot number is obtained from the cache slot management table in the shared control memory 112. If the I/O request is a read request and if the requested data exist in the cache memory 122, the data are retrieved from the cache memory 122 and sent to the host computer 100 through the loop 133 and HIFC 103. The presence or absence of the target data is determined by referencing the cache slot management table. If the requested data are not found in the cache memory 122, a write request is written to an inter-processor liaison area in the shared control memory 112. When the target data are judged to have been placed into the cache memory 122, the data are read from the cache memory 122 and sent to the host computer 100.

The control processors 119 and 120 search the cache slots in the cache memory 122 for any data to be written to the drives 127 through 130. If such data are detected, they are written to the drives 127 through 130 via the fibre channels 141, DIFCs 123 and 124, and loops 125 and 126. The write operation is carried out in a manner asynchronous with any I/O request processing between the host computer 100 on the one hand and the control processors 114 through 117 on the other hand. The control processors 119 and 120 convert the designated LUN and LBA into a physical device number and a physical address to determine the target drive and the address in the drive for the eventual write operation. The control processors 119 and

120 then reference the inter-processor liaison area in the shared control memory 112 to see if there is any data read request. If any such read request is found, the relevant data are read from the applicable drive or drives 127 through 130 and written to the relevant cache slot in the cache memory 122. Suitable entries are then updated in the cache slot management table to reflect the presence of the data.

I/O requests to the drives 127 through 130 may be processed by any one of the control processors 119 and 120. For example, if the control processor 119 or the fibre interface loop 125 has failed and is unusable, the processing of I/O requests is taken over by the control processor 120 and fibre interface loop 126. If either of the control processors fails, I/O request processing is carried out without interruption of I/O operations to and from the drives 127 through 130.

The control processors 114, 115, 116 and 117 monitor one another for operation status. Specifically, each processor writes the current time of day to the shared control memory 112 at predetermined intervals. The times posted by each processor are checked periodically by the other control processors for an elapsed time. If there is no difference between the preceding and the current time posting, the control processor in question is judged to have stopped. A control processor that has detected the stopped processor receives management information about the failed processor from the FCAL management information 113 and takes over the processing of the incapacitated processor. Illustratively, suppose that the control processor 114 has found the control processor 115 stopped. In that case, the control processor 114 updates the FCAL management information 113 as shown in Fig. 5. The updates allow the control processor 114 to take over the I/O requests regarding the LUNs 10-19 that had been processed by the control processor 115.

Each of the control processors 114 through 117 counts the number of

processed I/O requests and writes the counts to the shared control memory 112 at predetermined intervals. The control processors reference the processed request counts of one another to detect processors with inordinately high and low counts in order to average the counts therebetween. For example, suppose that the control processor 117 has found the control processor 116 with a falling processed request count and the control processor 115 with a rising request count. In that case, the control processor 117 updates the FCAL management information 113 as indicated in Fig. 6. It should be noted that relevant switch settings of the PBCs 108 through 111 need to be changed so that the frame with E8 in its AL-PA will be transmitted to the control processor 116 via the loop 133. The modifications allow the control processor 116 to process I/O requests with respect to the LUNs 10-19 and 20-29, whereby the processed request counts are averaged among the control processors to permit evenly distributed load processing.

Part of the LUNs 203 managed by a given control processor may be taken over by another control processor. For example, of the LUNs 10-19 managed by the control processor 115, solely the LUNs 15-19 may be taken over by the control processor 116. In that case, the FCAL management information 113 is updated as shown in Fig. 7. The control processors must inform the host computers 100, 101 and 102 of this change because the correspondence between the AL-PA 202 and LUN 203 is altered with regard to the LUNs 15-19.

The flow of processing by the control processors 114 through 117 has been described above with respect to the processing of I/O requests of the host computer 100 connected to the disk controller 107 via a fibre channel interface. Because the host computers 101 and 102 are connected to the disk controller 107 through interfaces different from the fibre channel interface, the HIFCs 104 and 105 convert

I/O commands received from the host computers 101 and 102 into a frame format compatible with the fibre channel interface before sending the converted commands to the control processors 114 through 117 via the loop 133. These arrangements make the processing of I/O requests sent from the host computers 101 and 102 equivalent to that which has been discussed above.

The HIFC 104 has functions for effecting conversion between commands, control information and data complying with an interface called ESCON (Enterprise System Connection) on the one hand, and commands, control information and data pursuant to the fibre channel interface on the other hand. The HIFC 105 is capable of providing conversion between commands, control information and data complying with the SCSI on the one hand, and commands, control information and data in keeping with the fibre channel interface on the other hand. When the disk controller 107 incorporates HIFCs having such host interface converting functions, any host computer may be connected to the disk controller 107 regardless of the type of host interface in use.

Although the embodiment above has been shown involving the drives 127 through 130 as disk drives, this is not limitative of the invention. Alternatively, magnetic tape units or floppy disk drives may be connected by modifying the DIFCs 123 and 124. If the DIFCs are equipped with functions for effecting conversion between the SCSI and the fibre channel interface, the loops 125 and 126 may be replaced by SCSI cables.

The disk controller 107 of this embodiment allows any one of the control processors 114 through 117 to handle I/O requests sent from the host computer 100. If a large number of I/O requests are coming from the host computer 100 depending on the data transfer rate between the computer 100 and the HIFC 103 or through the

loop 133, all of the control processors 114 through 117 can deal with the I/O requests. This provides a greater throughput than if fewer control processors were configured. Likewise, the I/O requests sent from the host computers 101 and 102 can be processed by any one of the control processors 114 through 117. When the host computers 100, 101 and 102 share the loop 133 and the control processors 114 through 117 in the manner described, it is possible for the inventive structure to have less lopsided load distribution among the components and ensure better performance of the storage controller as well as better cost/performance ratio than if the host computers 100, 101, 102, etc., have each an independent host interface connected to the common bus as in conventional setups.

As described, the storage controller according to the invention has its performance enhanced appreciably by having I/O requests from host computers processed in parallel by a plurality of control processors even as the processors have their loads distributed evenly therebetween. The invention is particularly conducive to making the most of high-speed fibre channel performance. The inventive storage controller is highly dependable because if any one of the control processors fails, the other processors take over the processing of the incapacitated processor.

The storage controller of the invention permits connection of multiple host computers having a plurality of kinds of interfaces, with the host computers sharing a fibre channel loop and control processors within the storage controller. This feature also promises excellent cost/performance ratio. Moreover, the storage controller permits connection of drives of different kinds of storage media.

As many apparently different embodiments of this invention may be made without departing from the spirit and scope thereof, it is to be understood that the

invention is not limited to the specific embodiments thereof except as defined in the appended claims.